DOCUMENT RESUME

ED 163 035 TM 008 036

AUTHOR Hymel, Glenn M.; Gaines, W. George
TITLE Emrick's Model as a Basis for Determining an Optimal Criterion Score and a Ratio of Regret in a Mastery Testing Situation.
PUB DATE 28 Mar 78
NOTE 15p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Toronto, Ontario, Canada, March 26, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Cost Effectiveness; Criterion Referenced Tests; *Cutting Scores; Item Analysis; *Mastery Tests; Mathematical Models; Test Reliability
IDENTIFIERS *Emrick (John A)

ABSTRACT

The evaluation model for mastery testing as proposed by Emrick (1971) represents an empirical approach to determining the most appropriate mastery criterion score that should be established in those testing situations which entail decision making relative to the continued progression or possible recycling of students in a given learning sequence. It is suggested in Emrick's model that practitioners and researchers alike are no longer limited to intuitive appeal or field-related precedents as a basis for establishing appropriate mastery score standards. Though not explicitly stated by Emrick, his model can be algebraically rewritten and its facility logically extended to accommodate those situations in which an ex post facto investigation is desired so as to determine the relative cost of decision errors associated with a previously administered mastery test which specified a given criterion score. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163035

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EMRICK'S MODEL AS A BASIS FOR DETERMINING AN OPTIMAL CRITERION SCORE AND A RATIO OF REGRET IN A MASTERY TESTING SITUATION

Glenn M. Hymel, Ed.D.
Loyola University
New Orleans, Louisiana

W. George Gaines, Ed.D.
Pro Data, Ltd.
Kenner, Louisiana

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Glenn M. Hymel

TO EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE ERIC SYSTEM."

Paper Presented
at the Annual Meeting of the
National Council on Measurement in Education

Toronto, Ontario, Canada
March 28, 1978

Printed in U.S.A.

TM008 036

ABSTRACT

There exists a paucity of research which addresses the most appropriate means for determining the optimal criterion score in a mastery testing situation. Emrick's model is an algorithm which generates a test cut rule or mastery criterion based upon both the properties of a given test and a cost-benefit analysis of possible decision errors. The data generated by the field-testing of a criterion-referenced formative test intended for use in a mastery learning setting is used to demonstrate the application of Emrick's model. Also, a possible ex post facto application of Emrick's model is suggested.

EMRICK'S MODEL AS A BASIS FOR DETERMINING AN OPTIMAL CRITERION SCORE AND A RATIO OF REGRET IN A MASTERY TESTING SITUATION

Background of the Problem

Proponents of criterion-referenced measurement and mastery or formative testing frequently couch their arguments within the context of a developmental view of the function of education. One important aspect of the developmental mode of education, as contrasted with the selective mode, asserts the necessity for providing alternative means of learning which are adaptive to an individual's background and current level of knowledge in a given subject area. The adaptive instructional models as developed by Glaser (1970) and Bloom (1968) are noteworthy examples of attempts to plan, organize, and implement pedagogical procedures in such a way that learner needs are accommodated and, hence, the maximum development of individuals is ensured.

Glaser & Nitko (1971), in a description of one of the integral components of any adaptive instructional model, make the following observations: As a student proceeds through a given course of instruction, his performance is continually monitored and assessed at specific test and decision points. The monitoring and assessment are based upon a pre-established criterion level of mastery which the student is expected to attain. Feedback from the mastery tests which are used for the monitoring and assessment purposes gives evidence as to whether the desired learning criterion has been met. In those instances where the achievement level is not satisfactory, instructional prescriptions of a review and/or remedial nature are provided.

The efficacy of the aforementioned feedback/correction approach to individualizing instruction has been demonstrated repeatedly by conclusive findings in the literature (Airasian, 1967; Arlin, 1973; Baley, 1972; Block, 1970, 1972, 1974; Block & Burns, 1976; Carroll & Spearritt; 1967; Collins, 1969, 1970;

Gentile, 1970; Gregware, 1974; Hymel, 1974; Keller, 1968; Kersh, 1970; Kim et al., 1969, 1970; Mayo, Hunt, & Tremmel, 1968; Merrill, Barton, & Wood, 1970; Moore, Mahan, & Ritts, 1968; Sherman, 1967; Silberman & Coulson, 1964). However, there presently exists a paucity of research which addresses itself specifically to the most appropriate means for determining how and at what percentage of proficiency a criterion level of mastery should be established in a given instructional setting.

Objectives of This Inquiry

Although the designation of 80-85% mastery is conventionally viewed as an appropriate criterion level in mastery testing situations, practitioners and researchers alike have suffered from an almost total lack of attention devoted to the formulation of a mathematical algorithm which generates a recommended mastery criterion based upon variables unique to a given instructional situation. The evaluation model for mastery testing as developed by Emrick (1971), though, appears to be a significant contribution in this area. With the preceding discussion in mind, then, the objectives of this paper are as follows:

1. To present the basic assumptions and formulation of Emrick's (1971) evaluation model for mastery testing.
2. To demonstrate the application of the model in determining the most appropriate mastery criterion that should be established for a field-tested achievement instrument.
3. To suggest an ex post facto application of the model to those situations in which the determination of a mastery criterion and the use of a given instrument have already occurred.

Although the major focus of this paper is on a suggested ex post facto application of Emrick's model, the authors pass no judgments relative to the correctness of his expression. The point is that Emrick's model--revised or

otherwise--can be algebraically rewritten and used in an ex post facto fashion and, thus, can permit practitioners to gain insight into the appropriateness of their decision making on previously administered mastery tests.

Emrick's Evaluation Model for Mastery Testing

John A. Emrick's (1971) evaluation model for mastery testing is an algorithm which generates a test cut rule or mastery criterion based upon both the properties of a given test and a cost-benefit analysis of possible decision errors. The five assumptions upon which the mastery test model is based have been described by Emrick in the following manner:

1. Relevant educational objectives are comprised of emplicitly defined skills. The mastery of each of these skills is of a binary nature, that is, all or none.
2. A given test designed to assess mastery of the component skills within an objective will consist of test items highly homogeneous relative to content, form, and difficulty level. Hence, the response to each item provides an unbiased estimate of the examinee's mastery status with respect to that skill.
3. Since the mastey of each skill is of a binary (all or none) nature, potential measurement error relative to a particular skill is only one of two types: (a) Type 1 or alpha (α) error--a "false pass," that is, a mastery conclusion when the examinee's true status is nonmastery; (b) Type 2 or beta (β) error--a "false fail," that is, a nonmastery conclusion when the examinee's true status is mastery.
4. The extent of measurement error in a skill test can be approximated by a determination of the average interitem correlation of examinee responses. This average interitem correlation, which provides an unbiased estimate of the squared correlation between an examinee's true mastery

state and his item response, can be interpreted in terms of the α and β measurement errors. Hence, $r = \frac{(1 - \alpha - \beta)^2}{1 - (\alpha - \beta)^2}$ is the expression for item reliability (r) on a single skill mastery test.

5. Due to the existence of at least some measurement error, decision errors relative to the examinee's status on the skill being measured will accrue. The regret associated with these evaluative or decision errors can be minimized through a cost-benefit analysis of the variables which comprise the evaluative process. These variables fall into statistical, curricular, and psychological classes and are related to regret stemming from decision errors in the following manner:

   a. As item reliability increases, tests of fixed length will yield proportionally fewer evaluative errors. Also, for a given item reliability, increasing test length functions to distinguish more clearly mastery from nonmastery examinees.

   b. Evaluative regret associated with instructional objectives of an ancillary nature would probably be low or irrelevant. However, evaluative regret associated with instructional objectives of a fundamental nature can become sizable. Concerning this latter point, the regret potentially associated with a Type 1 or "false pass" decision error could feasibly encompass costs of reduced efficiency in mastery of subsequent objectives due to nonmastery of prerequisites and, possibly, costs of diagnosis and remediation; regret potentially associated with a Type 2 or "false fail" decision error could perhaps encompass costs of unnecessary exercises, materials, instructional time, and retesting given the mastery student.

c. Evaluative regret associated with a Type 1 error could feasibly involve for the "out of track" learner psychological costs such as confusion and suboptimal success rate. Evaluative regret associated with a Type 2 error could perhaps encompass psychological costs such as boredom, decreased sense of achievement, and low motivation.

Emrick further contends that although meaningful quantitative values for each of the cost factors as well as the α and β item error probabilities are not conveniently available, a mastery criterion score for a given test can be optimized in terms of relative decision error costs and relative item error probabilities. The actual optimization formula and necessary elaboration as promulgated by Emrick (1971, p. 324) are as follows:

$$K = \frac{\log \frac{\beta}{1-\alpha} + \frac{\log RR}{n}}{\log \frac{\alpha\,\beta}{(1-\alpha)(1-\beta)}}$$

where K = the cut point expressed as a percent score on the test
α = estimated probability of Type 1 item error
β = estimated probability of Type 2 item error
RR = Ratio of Regret of Type 2 to Type 1 decision errors
n = test length (number of items).

. . . The total probability of item measurement error (α + β) is estimated as one minus the square root of the average interitem correlation (i.e., $1 - \sqrt{r}$). A logical analysis of item form yields an estimate of which type of error predominates. For example, a true-false test should yield relatively more α than β errors, whereas for recall or completion items, the reverse should be true.

A similar procedure is used to supply a value for the ratio of regret (RR). Specifically, a logical analysis of the evaluative procedure will yield an estimate of the more costly decision error, possibly in conjunction with some estimate of the examinee's status prior to testing. For example, in some testing situations, false fail errors may be considered far more costly than false passes whereas in other cases the reverse may be true.

An Application of Emrick's Model in Determining an Appropriate Mastery Criterion

The data generated by the field-testing of a criterion-referenced formative test intended for use in a mastery learning setting were used to demonstrate the application of Emrick's model. The instrument consisted of 25 algebraic problems which pertained to the addition, subtraction, multiplication, and division of directed numbers. A total of four objectives (related to the four operations, respectively) formed the basis for the actual construction of the test items. The instrument was administered to 70 male students at the secondary school level. The calculation of the instrument's internal consistency (via KR-20 formula) indicated a reliability coefficient of .86. This information, along with the derived values for the different variables in Emrick's formula, was used to determine empirically an optimal criterion score for this test. The sequential steps followed in this process and the corresponding results are summarized below:

1. The estimate of item reliability ($\bar{r}_i$), as determined by the use of the Spearman-Brown formula (Lord & Novick, 1968, p. 118), equaled .20.
2. Each of the 25 items necessitated some form of mathematical computation; the effects of guessing were therefore assumed to be negligible, thus suggesting a low α with respect to β --perhaps of a ratio of 1 to 10.
3. Based upon the simultaneous equations $\beta = 10\alpha$ and $(\alpha + \beta) = 1 - \sqrt{.20}$, α and β were estimated to have values of .0503 and .5025, respectively.
4. Since the instrument under consideration is a mastery test to be used in an instructional strategy involving feedback/correction procedures, it was decided that "false pass" decision errors would be considerably more detrimental than "false fail" decision errors. Hence, a ratio of regret equal to 1/10 was specified.

5. Given the above-cited values for the different variables in Emrick's formula, the optimal criterion score K was computed to be 46.97% of the total number of test items. The nearest item score to this value is 12 correct, or 48%.

A Suggested Ex Post Facto Application of Emrick's Model

As stated earlier, Emrick's model is intended for use prior to the actual implementation of a criterion-referenced mastery instrument in that it is designed to generate the most appropriate mastery criterion score for a particular test which possesses certain unique properties and which is intended for use in a certain type of testing situation. However, the model can be algebraically restated and its facility logically extended so as to make possible an ex post facto investigation of the ratio of regret that was operative in a situation where a criterion score was determined intuitively or by an appeal to field precedents and the instrument was then administered.

More specifically, given the original formulation of the model as

$$K = \frac{\log \frac{\beta}{1-\alpha} + \frac{\log RR}{n}}{\log \frac{\alpha\,\beta}{(1-\alpha)(1-\beta)}}, \qquad (1)$$

the following equation can be derived algebraically:

$$\log RR = n\left[\left(\log \frac{\alpha\,\beta}{(1-\alpha)(1-\beta)}\right)(K) - \log \frac{\beta}{(1-\alpha)}\right]. \qquad (2)$$

Although equation (2) is not discussed by Emrick, it is obviously implied in his original formulation and can be of definite service to practitioners and theoreticians alike. Though perhaps not labeled as Type 1 (false pass) or Type 2 (false fail) decision errors, it seems safe to assume that educators involved with mastery learning strategies would be aware of and sensitive to the two aforementioned kinds of error and, therefore, would be interested in deter-

mining in an ex post facto manner the relative weight that was associated with each type of decision error.

For the purpose of demonstrating the suggested ex post facto application of Emrick's model, let us assume that the instrument discussed earlier was indeed administered with a criterion level intuitively established at 60% mastery. This fact, along with the other test properties previously mentioned, results in the following findings relative to the ratio of regret:

1. log RR = -3.2
2. Hence, the ratio of regret (ratio of Type 2 decision error to Type 1 decision error) was approximately .001.

The situation just described for illustrative purposes gives evidence of a case in which the intuitive designation of a criterion score at the 60% level implied that "false pass" errors were tremendously more costly than "false fail" errors. This type of ratio of regret is exactly what one would be inclined to expect in a mastery testing situation in which the criterion score was set at such a level. It can be inferred, furthermore, that perhaps a moderate lowering of the required criterion level of mastery could have taken place with a correspondingly minimum increase in the ratio of regret which would still have preserved "false pass" decision errors as being more costly than "false fail" decision errors.

Summary

The evaluation model for mastery testing as proposed by Emrick (1971) represents an empirical approach to determining the most appropriate mastery criterion score that should be established in those testing situations which entail decision making relative to the continued progression or possible recycling of students in a given learning sequence. It is suggested in Emrick's model that practitioners and researchers alike are no longer limited to intuitive appeal

or field-related precedents as a basis for establishing appropriate mastery score standards. Though not explicitly stated by Emrick, his model can be algebraically rewritten and its facility logically extended to accommodate those situations in which an ex post facto investigation is desired so as to determine the relative cost of decision errors associated with a previously administered mastery test which specified a given criterion score.

References

Airasian, P. W. An application of a modified version of John Carroll's model of school learning. Unpublished master's thesis, University of Chicago, 1967.

Arlin, M. N. Rate and rate variance trends under mastery learning. Unpublished doctoral dissertation, University of Chicago, 1973.

Baley, J. D. Cost-effectiveness of three methods of remedial instruction in mastery learning and the relationship between aptitude and achievement. Unpublished doctoral dissertation, University of Southern California, 1972.

Block, J. H. The effects of various levels of performance on selected cognitive, affective, and time variables. Unpublished doctoral dissertation, University of Chicago, 1970.

Block, J. H. Mastery performance standards and student learning: A replication. Unpublished manuscript, Department of Education, University of California, Santa Barbara, 1972.

Block, J. H. Mastery learning in the classroom: An overview of recent research. In J. H. Block (Ed.), Schools, society and mastery learning. New York: Holt, Rinehart & Winston, Inc., 1974.

Block, J. H., & Burns, R. B. Mastery learning. In L. S. Shulman (Ed.), Review of research in education (Vol. 4). Itasca, Illinois: F. E. Peacock Publishers, Inc., and the American Educational Research Association, 1976.

Bloom, B. S. Learning for mastery. Evaluation Comment, 1968, 1, (2).

Carroll, J. B., & Spearritt, D. A study of a model of school learning. (Monograph No. 4) Cambridge, Mass., Harvard University, Center for Research and Development on Educational Differences, 1967.

Collins, K. M. A strategy for mastery learning in freshman mathematics. Unpublished study, Purdue University, Division of Mathematical Sciences, 1969.

Collins, K. M. A strategy for mastery learning in modern mathematics. Unpublished study, Purdue University, Division of Mathematical Sciences, 1970.

Emrick, J. A. An evaluation model for mastery testing. Journal of Educational Measurement. 1971, 8, 321-326.

Gentile, J. R. A mastery strategy for introductory educational psychology. Unpublished materials, State University of New York at Buffalo, Department of Educational Psychology, 1970.

Glaser, R. Evaluation of instructon and changing educational models. In M. C. Wittrock & D. E. Wiley (Eds.), Evaluation of instruction. New York: Holt, Rinehart, & Winston, 1970.

Glaser, R. & Nitko, A. J. Measurement in learning and instruction. In R. L. Thorndike (Ed.), Educational measurement, 2nd ed. Washington, D. C.: American Council on Education, 1971.

Gregware, G. The interactions of three variables in John B. Carroll's model of school learning. Unpublished master's thesis, University of New Orleans, 1974.

Hymel, G. M. An investigation of John B. Carroll's model of school learning as a theoretical basis for the organizational structuring of schools. Published doctoral dissertation, University of New Orleans, 1974.

Keller, F. S. Goodbye, teacher . . . . Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Kersh, M. E. A strategy for mastery learning in fifth-grade arithmetic. Unpublished doctoral dissertation, University of Chicago, 1970.

Kim, H., et al. A study of the Bloom strategy for mastery learning. Seoul: Korean Institute for Research in the Behavioral Sciences, 1969. Cited by J. H. Block (Ed.), Mastery learning: Theory and practice. New York: Holt, Rinehart & Winston, Inc., 1971.

Lork, F. M., & Novick, M. R. Statistical theories of mental test scores. Massachusetts: Addison-Wesley, 1968.

Mayo, S. T., Hunt, R. C., & Tremmel, F. A mastery approach to the evaluation of learning statistics. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, 1968.

Merrill, M. D., Barton, K., & Wood, L. E. Specific review in learning a hierarchical imaginary science. Journal of Educational Psychology, 1970, 61, 102-109.

Moore, J. W., Mahan, J. M., & Ritts, C. A. An evaluation of the continuous progress concept of instruction with university students. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1968.

Sherman, J. G. Application of reinforcement principles to a college course. Paper presented at the annual meeting of the American Educational Research Association, New York, 1967.

Silberman, H., & Coulson, J. Final report: Use of exploratory research and individual tutoring techniques for the development of programing methods and theory. Santa Monica, Calif.: System Development Corporation, 1964.